Article

Prediction of Diabetic Nephropathy from the Relationship between Fatigue, Sleep and Quality of Life

Angela Shin-Yu Lien [1,2,3] [iD], **Yi-Der Jiang** [4] [iD], **Jia-Ling Tsai** [1,5], **Jawl-Shan Hwang** [3] and **Wei-Chao Lin** [2,6,7,*] [iD]

[1] School of Nursing, College of Medicine, Chang Gung University, Taoyuan 333, Taiwan; angela@mail.cgu.edu.tw (A.S.-Y.L.); jialing5147@gmail.com (J.-L.T.)
[2] Healthy Aging Research Center, Chang Gung University, Taoyuan 333, Taiwan
[3] Division of Endocrinology and Metabolism, Department of Internal Medicine, Chang Gung Memorial Hospital, Linkou 333, Taiwan; hwang2570@gmail.com
[4] Department of Internal Medicine, National Taiwan University Hospital, Taipei 100, Taiwan; yiderj@ntu.edu.tw
[5] School of Nursing, College of Medicine, National Taiwan University, Taipei 106, Taiwan
[6] Department of Information Management, Chang Gung University, Taoyuan 333, Taiwan
[7] Department of Thoracic Surgery, Chang Gung Memorial Hospital, Linkou 333, Taiwan
[*] Correspondence: viclin@gap.cgu.edu.tw; Tel.: +886-3-211-8800 (ext. 3770)

check for updates

Abstract: Fatigue and poor sleep quality are the most common clinical complaints of people with diabetes mellitus (DM). These complaints are early signs of DM and are closely related to diabetic control and the presence of complications, which lead to a decline in the quality of life. Therefore, an accurate measurement of the relationship between fatigue, sleep status, and the complication of DM nephropathy could lead to a specific definition of fatigue and an appropriate medical treatment. This study recruited 307 people with Type 2 diabetes from two medical centers in Northern Taiwan through a questionnaire survey and a retrospective investigation of medical records. In an attempt to identify the related factors and accurately predict diabetic nephropathy, we applied hybrid research methods, integrated biostatistics, and feature selection methods in data mining and machine learning to compare and verify the results. Consequently, the results demonstrated that patients with diabetic nephropathy have a higher fatigue level and Charlson comorbidity index (CCI) score than without neuropathy, the presence of neuropathy leads to poor sleep quality, lower quality of life, and poor metabolism. Furthermore, by considering feature selection in selecting representative features or variables, we achieved consistence results with a support vector machine (SVM) classifier and merely ten representative factors and a prediction accuracy as high as 74% in predicting the presence of diabetic nephropathy.

Keywords: fatigue; sleep quality; quality of life; diabetic nephropathy; feature selection; data mining

1. Introduction

Diabetes mellitus (DM) is defined as a series of metabolic disorders induced by insulin insufficiency or the inability to use insulin. Insulin is a hormone secreted by the pancreas that brings glucose from the bloodstream into the cell body after we consume food. When our body is not able to use insulin, hyperglycemia (rise in the blood glucose level) occurs, causing diabetes and related complications. Diabetes and its comorbidities plague more than a quarter of the world's population. Diabetic patients often experience fatigue and have sleep disorders. Given that fatigue and poor sleep quality are

subjective and unquantifiable, it is difficult to define it specifically and treat it medically, but it truly affects patients' quality of life (QoL) and behavioral ability. According to the literature, diabetes-related fatigue can be caused by three major categories of factors: first, physiological factors, including acute or chronic hyperglycemia and hypoglycemia, glycemic variability, and diabetes symptoms; second, psychological factors, including emotional stress and sleep problems; third, lifestyle factors, including increase in weight or body mass index and decrease in physical activity [1]. Thus, fatigue is not entirely psychological and is usually accompanied by actual physical weakening. The signs that indicate that fatigue may be derived from physical weakening include muscle weakness, a lack of spirit, weight loss, sluggishness, and reduced vitality [2,3]. However, in the early stage of the comorbidity, patients with diabetes are also likely to suffer from a decline in sleep quality; hence, an accurate quantitative assessment of lifestyle, exhaustion, and sleep status of diabetic patients is not only precautionary but necessary.

Moreover, poor control of DM induces multiple complications, such as vascular disease, neuropathy, eye lesions, and renal disease, and long-term medical attention has become an encumbrance worldwide. Among all the complications, diabetic nephropathy is associated with the highest costs for individuals, families, and society. Due to the fact that nephropathy aggravates extreme fatigue and insomnia especially when patients need dialysis [4–6], recent studies have noted that insomnia, fatigue, and physical weakness symptoms in diabetic patients are found not only in elderly patients, but also in adolescents [7] and adults [8,9], and a debilitating condition in individuals who are not yet middle-aged is indicative of early aging [10]. Because fatigue is closely related to sleep status and glycemic control, it has been reported that a reduction in the workplace performance of fatigued diabetic patients is indirectly linked to depression [11]. It is important to note that a lack of sleep often results in a lower spirit or even depression, especially for people with diabetes. [1,12]. Moreover, only traditional statistical methods including regression models have been used, and the multiple complex relationships between fatigue, sleep, physiological status, and lifestyles in diabetic patients still need to be verified [13]. Therefore, the development of an accurate and reliable method for predicting the degree of fatigue and sleep quality of patients with complications will be important for determining the effectiveness of the disease control and improving quality of life.

Based on the above rationale, the purpose of this study was to clarify the complex relationships between the presences of diabetes complications and sleep quality, fatigue, and metabolic control. To explore relevant factors and construct a better prediction model for diabetic nephropathy, research results suggest using a hybrid method by combining traditional statistics and machine learning methods to manage the large amount of data concerning all aspects of diabetes (such as diagnostic criteria, lifestyle factors, and treatment-related factors) [14,15]. The use of an adequate feature-selection method can enhance the accuracy of the classification results to support a better conclusion model [16]. Therefore, this study utilized chi-square verification, regression, logistic regression, and feature selection methods in data mining and machine learning for a novel hybrid methodology.

A literature review is presented in Section 2 to discuss the current research on fatigue, sleep, and quality of life in diabetic patients. The methodology, including the study design, subjects, data collection procedure, and analysis, is described in Section 3. The findings of the two analyses are addressed in Section 4. The innovative aspects and conclusions are discussed in the final section.

2. Literature Review

2.1. Diabetes-Related Fatigue and Sleep Disturbance

Studies have shown that people with diabetes often suffer from nighttime hypoglycemia due to an inadequate diet at dinner and an insulin injection overdose, which can cause problems including nightmares, sleep disorders, or respiratory distress [17]. Furthermore, depression caused by sleep disorders, restless legs, and sleeplessness can affect the patient's metabolic control and reflect the treatment effectiveness for diabetes on the glycosylated hemoglobin (HbA1c) level [18]. Since the

HbA1c level reflects the average blood glucose level of patients in the previous three months, it is the most accurate quality indicator of diabetic control and predictor of the presence of complications; multiple complications lead to an extreme medical burden of costs and a higher mortality rate for patients with diabetes. Therefore, preventing complication has always been the most critical component in treating diabetes [19]. Regardless, diabetic nephropathy affects the health of one-third of patients who have Type 1 or Type 2 diabetes. [20]. Some clinical signs and symptoms of poor diabetic control are manifested early. The most common complaint and an early sign is fatigue or tiredness and poor sleep quality. However, a painful sensation is a complaint that is reported earlier than fatigue in the presence of complications [1]. Diabetes fatigue syndrome is defined as a multifactorial syndrome of fatigue or easy fatigability, occurring in persons with diabetes [21]. The conclusion from a randomized clinical trial (RCT) in the United States was that multidimensional phenomena (physiological, psychological, and lifestyle phenomena) are associated with fatigue in diabetes patients [22], so clarifying the relationship is important to improve the quality of care. Furthermore, Fritschi and Quinn [1] noted that when diabetes leads to fatigue, good self-care and glycemic control can possibly improve the level of fatigue and quality of life of patients. However, due to the lack of a consistent definition, the research on diabetes-related fatigue is inadequate and lacks specificity; we strongly encourage researchers to further clarify the relationship between related factors by assigning research participants randomly.

2.2. The Presence of the DM Complication Nephropathy

According to a global estimate, the worldwide incidence of diabetic nephropathy is expected to reach a highest level by 2050. This will be associated with a large medical expense and economic burden due to DM complications [23]. The complications, such as nephropathy, have been suggested to be related to diabetes fatigue syndrome. DM nephropathy means the kidney function been damaged by long-term poor metabolic control. It will reduce the function of removing waste products and extra fluid from our body. If the patient has pre-existing symptoms or diseases, for example, hypertension, or cardiovascular disease, the risk of end-stage renal disease (ESRD) will increase. Diabetes is the leading cause of the end-stage renal disease (ESRD), and individuals with ESRD require dialysis. The best way to prevent or delay diabetic nephropathy is by maintaining a healthy lifestyle and keeping alert of the symptom of poor diabetic management and early signs and symptoms of complications. The clinical community has begun to define chronic fatigue caused by diabetes in detail [24] and explore the relationship between the symptoms and complications [25], allowing diabetic patients to have better symptom control to improve their quality of life and prevent complications. Therefore, the purpose of this study was to investigate the occurrence of diabetic nephropathy and its indicators. The results of this study complement the current knowledge in the literature, guiding healthcare professionals in improving patient care.

2.3. Feature Selection

Feature selection (or variable selection) is a very important data preprocessing step in data mining and pattern recognition aimed at filtering out unrepresentative features or selecting a subset of features from a given training dataset. There are several advantages of performing feature selection. Since higher feature dimensions are likely to have noisier features, the curse of the dimensionality problem can be minimized. Moreover, once a certain number of noisy features are removed, the overfitting problem and training time for model construction can be reduced. Finally, the performance of the constructed models based on the feature subset could be better than the ones based on the original feature set [26–29].

In general, feature selection methods can be classified into three categories: filter, wrapper, and embedded (or hybrid) [26,28,30–32].

2.3.1. Filter Methods

One major type of filter method to select important features is based on a ranking technique. Specifically, the input features are scored via a suitable ranking criterion, and the features that have a rank below a certain threshold are removed.

Filter methods usually require the following procedures. For methods based on a particular search strategy, a given subset, which may be an empty set, a full set, or a randomly-selected subset from a given dataset, is searched. Then, each generated subset is evaluated by a specific measure and compared with the previous best subset. This search process iterates until the predefined stopping criterion is met. Consequently, the final output of this method is the most recent best subset.

Many statistical techniques are used in filter methods, including principal component analysis, information gain, and stepwise regression. This type of technique has excellent computational efficiency and independent classification/clustering algorithms.

For example, for principal component analysis (PCA), the central idea is to reduce the dimensionality of a data set in which there are a large number of interrelated variables while retaining as much variation as possible in the dataset. This reduction is achieved by transforming the variables to a new set of variables, such as the principal components, which are uncorrelated and ordered so that the first few components retain most of the variation present in the entire set of original variables. By computing the eigenvalues and eigenvectors of the principal components, we can identify a linear combination of the original variables that explains the most variance. The first principal component accounts for the highest amount of the variability in the data as possible, and each successive component accounts for the highest amount of the remaining variability as possible. Thus, the definition and computation of principal components are straightforward [31].

On the other hand, partial least squares regression (PLS), which relates to principal component regression, focuses on modelling the covariance structures in the spaces of predicted variables and the observable variables. In other words, it finds a linear regression model by projecting the predicted variables and the observable variables to a new space.

2.3.2. Wrapper Methods

The wrapper methods are based on using a predictor (or learning model) as the objective function to evaluate different feature subsets. The subset that allows the predictor to generate the highest accuracy rate is chosen as the best feature subset. In other words, the wrapper-based feature selection methods are similar to the filter-based methods, except for the inclusion of a predefined mining algorithm, which is utilized as the search strategy and evaluation measure. Therefore, using different objective functions will produce different feature selection results with the same dataset.

Evolutionary computational techniques, such as the genetic algorithm and particle swarm optimization, have recently gained attention and have shown some success [33–35]. In addition, the genetic algorithm and particle swarm optimization have been regarded as representative wrapper methods. However, they entail a large computational cost for model training and searching for the best subset.

For example, with the genetic algorithm, a population of strings (called chromosomes), in which candidate solutions (called individuals) to an optimization problem are encoded, evolve to identify better solutions. In general, the genetic information (i.e., chromosome) is represented by a bit string (such as a binary string of 0 s and 1 s), and the solution is encoded in sets of bits. Then, genetic operators are applied to the individuals of the population in the next generation (i.e., a new population of individuals). In the crossover process, two offspring strings are created from two parent strings when selected bits from each parent are copied. On the other hand, in the mutation process, the value of a single bit (with small probability) within the bit strings randomly changes. Furthermore, a fitness function is used to measure the quality of an individual to increase the probability that the single bit can survive throughout the evolutionary process [36]. As a result, a bit value of 1 in the final chromosome

representation indicates that the corresponding feature is selected in the subset, whereas a value of 0 indicates otherwise.

2.3.3. Embedded Methods

With embedded methods, feature selection is performed during the model learning process [37–39]. In other words, feature selection is incorporated into the classifier training process. Specifically, embedded methods not only measure the relations between the input features and the output features but also search for features that allow better classification accuracy. One representative embedded method is the decision tree model, where the constructed tree contains a number of selected features (i.e., decision nodes) that can distinguish between different classes (i.e., leaf nodes) well. In addition to decision trees, there are some other types of embedded feature selection methods, such as l_1-regularization techniques, including LASSO (least absolute shrinkage and selection operator) [40], l_1-SVM (support vector machine) [41], and memetic algorithms [42].

For example, with C4.5, a tree-like graph or model of decisions is constructed from a given dataset. At each node of the tree, the attribute of the dataset that most effectively splits its set of samples into homogeneous classes is selected. In particular, the splitting criterion is based on normalized information gain, so the attribute with the highest gain is chosen to make the decision [43]. Therefore, the nodes of the tree can be regarded as the most representative attributes (or features) of the dataset for classification purposes.

3. Methodology

3.1. Data Collection

Based on a questionnaire survey and retrospective study on medical records of 307 adults with Type 2 diabetes in the Department of Endocrinology and Metabolism of two medical centers in Northern Taiwan, this study was designed to investigate the relationship between fatigue, sleep quality, quality of life, and the presence of diabetic nephropathy.

3.2. Instruments

Questionnaires and scales were chosen for hypothesis testing in this study based on the literature review. Patients with type 2 diabetes (T2DM) should receive lifestyle modifications under a regular medication treatment and a metabolic follow-up to prevent the occurrence of multiple complications that induce poor quality of life [4–6]. Therefore, we collected basic lifestyle information, physical status, sleep quality measurements (PSQI), multidimensional fatigue inventory (MFI) results, and quality of life SF-12 questionnaire results. Furthermore, the metabolic control index was included and determined from blood tests and urine analyses, so 128 features were assessed in this study. The data collection process and the features were assessed from the associated factors are shown in Figure 1.
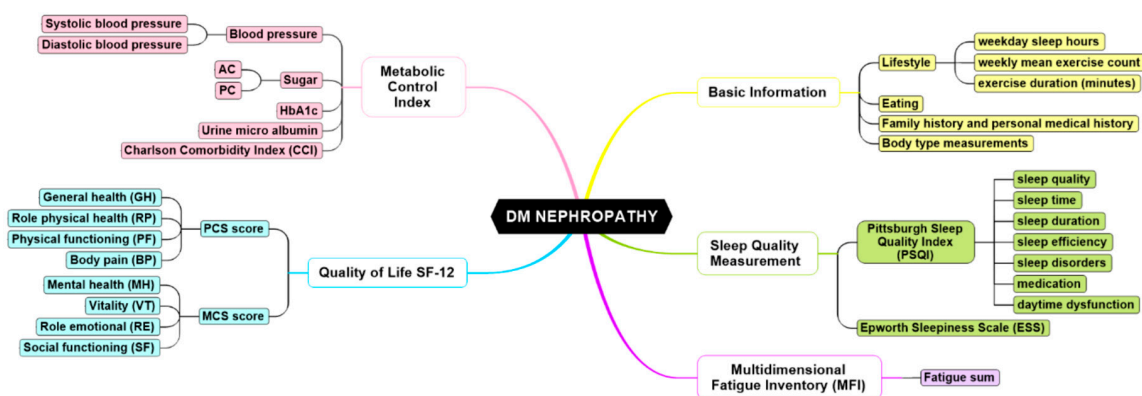

Figure 1. The data collection process and the features obtained from the questionnaires.

Questionnaire one: Basic information and lifestyles

The questionnaire consisted of 25 questions, which surveyed the respondents' living habits in the past six months, including (1) lifestyle habits (smoking, drinking, betel nut chewing, and exercise habits); (2) eating habits (red meat, fish, fruit, and vegetable consumption, and dining out frequency); (3) family history and personal medical history; and (4) body type measurements (height, weight, body mass index, waist circumference, and hip circumference).

Questionnaire two: Sleep quality measurement

The Pittsburgh sleep quality index (PSQI) is a self-report questionnaire that measures sleep quality and disorders and is widely used clinically, and it has superior reliability and validity for patients with a history of depression and sleep disorders. The scale includes seven components concerning the previous month, including subjective sleep quality, sleep time, sleep duration, sleep efficiency, sleep disorders, medication, and daytime dysfunction, and the scores of each component can be summed differently with the use of different rules. The reliability as measured by Cronbach's alpha was 0.72.

Questionnaire three: Multidimensional fatigue inventory (MFI)

A total of 20 questions were used to explore the subjective fatigue of the patients and the impact on their lives. The reliability as measured by Cronbach's alpha was 0.88.

Questionnaire four: Quality of life (QoL) SF-12

The SF-12 questionnaire designed by Ware et al. was used, and the reliability as measured by Cronbach's alpha was 0.81 for quality of life in this study.

The second part included a retrospective study on the metabolic control index and the diagnosis of diabetic nephropathy. The indicators include systolic and diastolic blood pressure, AC sugar (fasting plasma glucose level) and PC sugar (postprandial 2 hours' plasma glucose level), HbA1C (glycosylated hemoglobin), urine micro albumin, and the Charlson comorbidity index (CCI), and the results were obtained retrospectively from the medical records on the day of the patient's visit. The main purpose of this part was to determine whether renal disease is related to fatigue, sleep status, and quality of life in diabetic patients and the relationship between these factors.

4. Experiment Results

4.1. Statistical Analysis

This study investigated the relationship between fatigue and sleep status, diabetes control, quality of life, and diabetic nephropathy in diabetic patients. Continuous variables, such as age, waist circumference, and blood pressure, are presented as descriptive statistics, including the percentage, mean, and standard deviation. To compare the continuous variables and categorical variables, chi-squares tests and unpaired two-sample t-tests were used. Considering the factors (age, sex) that influence the result, univariate and multivariate analyses were also used to explore the relationships between fatigue status and various characteristics of the patients, and the relationships between the relevant attributes were analyzed by correlation tests, linear regression, or logistic regression. A two-sided *p*-value less than 5% was considered statistically significant. All analyses were conducted with SAS version 9.3 (SAS® Institute, Cary, NC, USA).

4.2. Results of Statistical Analysis

As shown in Table 1, 89 of the 307 patients (28.9%) were diagnosed with nephropathy; the mean age was 63.49 ± 10.42, and there were no significant differences in age between the patients with and without nephropathy. Males were found to be twice as likely to have renal lesions as females, with a significant difference between the sexes. There were no significant differences in the metabolic control index (AC, PC sugar, HbA1c, etc.,) among the patients; however, patients with renal disease showed

worse physical measurements (waist circumference, hip circumference, waist-hip ratio, BMI, systolic blood pressure, and diastolic blood pressure) than those without renal disease, and the differences were significant. Additionally, the presence of a comorbidity is also an important indicator. The comorbidity score of the patients with nephropathy determined by age-adjusted CCI was 6.754 ± 1.26, which was much higher than that of the patients without nephropathy. This result shows that when renal lesions appear, the patient is likely to simultaneously have comorbid conditions ($p = 0.000$). In terms of lifestyle, the diabetic patients without renal disease tended to have a higher level of average weekly exercise and less fatigue than those without renal disease, and they showed significantly better results in both the role physical health (RP) and physical function (RF) domains. Although there were no significant differences in sleep status (PSQI) or sleepiness (ESS) between the two groups according to the chi-squared test, both groups of patients had a worse sleep status and lethargy.

Table 1. Characteristics of the study patients with and without nephropathy ($n = 307$).

Variables	Total	Nephropathy		p-Value
		Yes	No	
	($n = 307$)	($n = 89$)	($n = 218$)	p-Value
	Mean ± SD	Mean ± SD	Mean ± SD	
Age (y)	63.49 ± 10.42	64.63 ± 11.19	63.02 ± 10.07	0.219
Physical status				
Waist circumference	89.10 ± 11.12	94.58 ± 11.49	86.87 ± 10.18	0.000 ***
Hip circumference	97.63 ± 8.56	100.68 ± 8.88	96.41 ± 8.13	0.000 ***
Waist/hip ratio	91.16 ± 0.07	0.94 ± 0.76	0.90 ± 0.61	0.000 ***
BMI	25.15 ± 4.20	26.44 ± 4.68	24.62 ± 3.88	0.001 **
Systolic blood pressure	128.49 ± 13.43	131.69 ± 13.49	127.04 ± 13.17	0.004 **
Diastolic blood pressure	75.53 ± 9.33	77.97 ± 9.03	74.51 ± 9.29	0.003 *
Metabolic control index				
AC sugar (preprandial sugar)	135.13 ± 37.16	137.08 ± 46.20	134.33 ± 32.83	0.611
PC sugar (postprandial sugar)	195.02 ± 67.64	202.48 ± 61.05	191.98 ± 70.10	0.296
Glycosylated hemoglobin (HbA1c)	7.60 ± 3.86	8.30 ± 7.01	7.32 ± 0.97	0.197
Complications				
Age-adjusted Charlson comorbidity index (CCI)	4.54 ± 1.87	6.75 ± 1.26	3.63 ± 1.21	0.000 ***
Lifestyle				
Weekday sleep hours	7.41 ± 2.29	2.13 ± 0.14	2.63 ± 0.28	0.113
Weekly mean exercise count	3.68 ± 2.67	2.55 ± 0.17	2.90 ± 0.31	0.047 *
Exercise duration (minutes)	45.91 ± 40.26	39.41 ± 2.67	40.98 ± 4.34	0.008 *
Total exercise minutes/week	207.25 ± 218.79	215.91 ± 14.62	223.74 ± 23.72	0.104
Fatigue sum	52.36 ± 9.00	54.2 ± 10.04	51.6 ± 8.45	0.021 *
Sleepiness (ESS) score	8.77 ± 4.92	8.84 ± 5.53	8.74 ± 4.66	0.876
sleep status (PSQI) score	7.41 ± 3.91	7.48 ± 3.68	7.39 ± 4.00	0.837
Quality of Life (SF-12) score	70.56 ± 13.92	68.17 ± 16.82	71.54 ± 12.46	0.090
PCS score	68.56 ± 18.53	65.10 ± 21.77	69.97 ± 16.83	0.062
General health (GH)	35.34 ± 21.12	35.67 ± 23.50	35.21 ± 20.13	0.896
Role physical health (RP)	80.46 ± 27.87	74.72 ± 31.49	82.80 ± 26.01	0.034 *
Physical functioning (RF)	73.98 ± 25.40	31.42 ± 30.80	77.01 ± 22.37	0.004 *
Body pain (BP)	84.45 ± 21.49	83.43 ± 22.60	84.86 ± 21.07	0.596
MCS score	72.57 ± 12.95	71.54 ± 14.89	73.11 ± 12.09	0.253
Mental health (MH)	77.48 ± 22.88	74.30 ± 26.81	78.78 ± 20.99	0.160
Vitality (VT)	77.20 ± 18.67	75.56 ± 18.84	77.87 ± 12.60	0.327
Role emotional (RE)	55.46 ± 12.41	57.58 ± 12.67	54.59 ± 11.78	0.055
Social functioning (SF)	80.13 ± 23.49	77.53 ± 26.66	81.19 ± 22.04	0.254

* $p < 0.05$, ** $p < 0.001$, *** $p < 0.0001$.

In Table 2, we further discussed the relationships between fatigue, drowsiness, sleep status, quality of life, glycemic control indicators, and CCI in patients with diabetes. It was found that the degree of fatigue, sleepiness, sleep quality, and quality of life are closely and significantly related to the comorbidities. The fatigue score was highly correlated with the presence of sleepiness symptoms, especially in the daytime, and poor bedtime sleep quality. We also found that the fatigue scores were

negatively correlated with patients' self-awareness of their quality of life, indicating that a higher fatigue score corresponds to a lower quality of life for patients with diabetes. Moreover, the fatigue score was highly correlated with the severity of the comorbidities with statistical significance. However, we did not find a correlation between glucose control (HbA1c) and fatigue, sleepiness, and sleep quality and comorbidities.

Table 2. Correlation between fatigue, ESS, PSQI, SF-12, HbA1C, and comorbidities ($n = 307$).

	Fatigue Score	ESS Score	PSQI Score	SF-12 Score	HbA1C	CCI
	r (p)	*r (p)*	*r (p)*	*r (p)*	*r (p)*	*r (p)*
Fatigue score	1.000					
ESS score	0.192 (0.001 **)	1.000				
PSQI score	0.395 (0.000 ***)	0.060 (0.292)	1.000			
SF-12 score	−0.605 (0.000 ***)	−0.299 (0.000 ***)	−0.342 (0.000 ***)	1.000		
HbA1c	0.029 0.612	−0.012 (0.840)	0.012 (0.839)	0.026 (0.647)	1.000	
CCI	0.221 (0.000 ***)	0.021 (0.714)	0.067 (0.245)	−0.200 (0.000 ***)	0.068 (0.238)	1.000

* $p < 0.05$, ** $p < 0.001$, *** $p < 0.0001$.

As seen in Table 3, we used a logistic regression method and the odds ratio to examine whether the presence of nephropathy can be predicted by sleep quality, sleepiness, quality of life, disease control, and the comorbidity score in DM patients. After adjusting for the variables, when HbA1c was high, the incidence of renal lesions increased by 1.776 times (aOR: 1.776, 95%; CI: 1.186–2.660, $p = 0.005$), and the comorbidity index also increased by 9.777 times (aOR: 9.776, 95%; CI: 5.186–17.591, $p < 0.000$). Logistic regression analysis was used for all the variables (HbA1c, fatigue, sleep status, lethargy, etc.). A high HbA1c indicates poor glucose control, which is an important predictive factor of comorbidities, especially DM nephropathy. Furthermore, a high comorbidity index score is predictive of an increased risk of DM nephropathy.

Table 3. Logistic regression analysis of nephropathy risk factors with fatigue, ESS, PSQI, SF-12, HbA1c, and CCI included as continuous variables.

	Univariate			Multivariate #		
	OR	95% CI	*p*	aOR	95% CI	*p*
Fatigue score	1.033	(1.004–1.062)	0.023 *	0.979	(0.917–1.045)	0.524
ESS score	1.004	(0.955–1.056)	0.866	1.010	(0.916–1.112)	0.847
PSQI score	1.006	(0.945–1.072)	0.842	0.898	(0.782–1.031)	0.126
SF-12 score	0.983	(0.966–1.000)	0.056	0.995	(0.952–1.041)	0.837
HbA1c	1.200	(0.970–1.484)	0.094	1.776	(1.186–2.660)	0.005 **
CCI	7.779	(4.749–12.742)	0.000 ***	9.777	(5.435–17.591)	0.000 ***

Multivariate analysis has adjusted for fatigue, ESS, PSQI, SF-12, HbA1c, and CCI, * $p < 0.05$, ** $p < 0.001$, *** $p < 0.0001$.

4.3. Feature Selection Analysis

Here, principal component analysis (PCA) [44,45], the genetic algorithm (GA) [33,34], and the C4.5 decision tree [46,47] are used as the filter, wrapper, and embedded methods, respectively. In related literature, these three feature selection methods have been widely used as representatives of the three types of feature selection methods [48–50].

For the collected dataset containing 182 features, using different feature selection methods could result in different numbers of features being selected. To evaluate the representativeness of the selected features from different feature selection methods, the predication accuracy of the support vector machine (SVM) [51] was examined. That is, the selected features that allow the SVM to output the highest rate of prediction accuracy in distinguishing patients with and without neuropathy can be regarded as the most representative features.

4.4. Results of the Feature Selection Analysis

Figure 2 shows the prediction accuracy of the SVM with and without feature selection. Note that the 5-fold cross validation method was used to divide the dataset into 80% training and 20% testing sets in order to construct the SVM classifier and examine its prediction accuracy, respectively. That is, the baseline approach was based on an SVM without feature selection, and three related feature selection methods were combined with an SVM individually: GA, PCA, and C4.5. The features selected by C4.5 allowed the SVM to provide the highest rate of prediction accuracy.
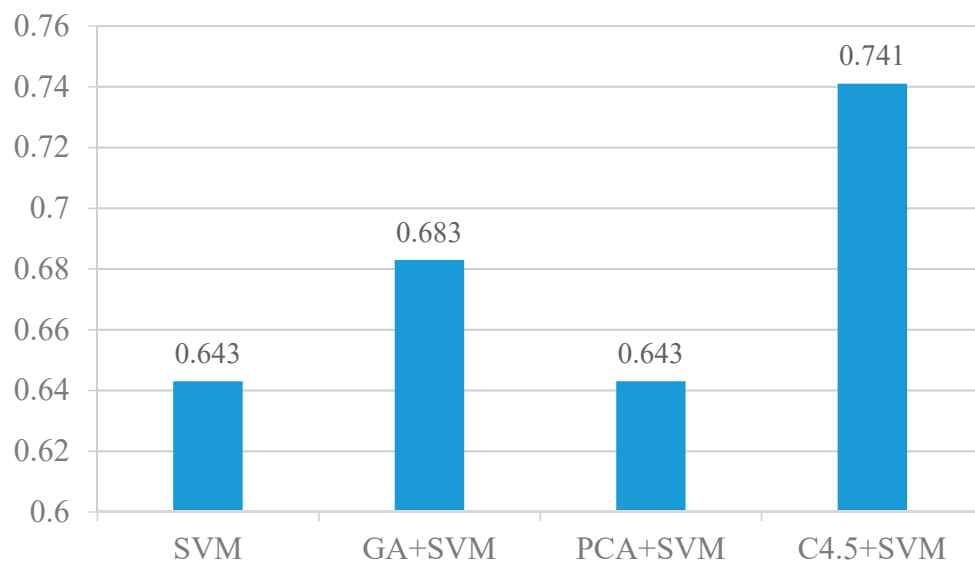
Figure 2. Prediction accuracy of SVM (support vector machine), PCA (principal component analysis) + SVM, GA (genetic algorithm) + SVM, and C4.5(C4.5 decision tree) + SVM.

Specifically, the GA, PCA, and C4.5 selected 140, 60, and 10 features from the 182 features, respectively. It is surprisingly effective to use only 10 features selected by the C4.5 to distinguish patients with and without nephropathy. Table 4 lists the 10 representative features selected by the C4.5 method.

Table 4. The 10 representative variables.

Variable Names
PC sugar (postprandial 2 hours' plasma glucose level), urine microalbumin, microalbumin-to-creatinine ratio, body height, body weight, waist circumference, hip circumference, diastolic blood pressure, alcohol drinking behavior, exercise.

5. Discussion and Conclusions

This is the first study to use a hybrid method, which includes traditional statistics and data-mining methods, to effectively predict the presence of diabetic nephropathy from the relationships between fatigue, sleep, and quality of life. The results show the correlation between fatigue, sleep quality, quality of life, and renal disease in diabetic patients. Using traditional statistics, we found significant correlations between related factors and diabetic nephropathy. Moreover, patients with renal disease had physical indicators (BMI, WC, HC, waist-to-hip ratio, and BP) and lifestyle factors that were significantly different from those without renal disease, and they had more severe comorbidities and were more fatigued. The results showed that when diabetic patients have poor disease control and comorbidities occur, fatigue, lethargy, and sleep disorders also appear, which confirms the conclusions in a previous study [52]. Moreover, this study also found that patients with renal syndrome had poor role physical health (PH) and physical function (PF) and lower amounts of exercise, as well as a

higher comorbidity index. Regression models were used to determine the indicators (fatigue scores, HbA1c, and comorbidities) that can predict diabetic nephropathy. This result is consistent with those in previous studies that concluded uncontrolled glycemia is highly associated with fatigue [21,53].

In addition, we used a feature selection method to determine the best predictive model by applying only 10 representative metrics (PC sugar, urine microalbumin, M/C, BH, BW, WC, HC, DBP, alcohol drinking behavior, and exercise). In other words, with the application of these representative metrics, the SVM classifier predicted the presence of diabetic nephropathy with an accuracy as high as 74%. Furthermore, alcohol drinking behavior was not identified by traditional statistical models but was instead selected by the SVM model for predicting the presence of complications in patients with diabetic nephropathy. Therefore, the hybrid method enhances the accuracy and validity of the prediction model.

On the basis of the results of this study, when a patient experiences lethargy, reduced physical strength, and insomnia, one should carefully examine whether the patient already has diabetes complications that are in the initial stage. Muscle strength and stamina training; proper nutrition, especially vitamin D and high quality protein intake; routine preventive urine screenings; and appropriate metabolic control can not only improve physical function but also slow the deterioration of the disease [54,55]. Finally, to provide better quality of care, we suggest using a hybrid method to make a precise predication model to explore the related risk factors and the relationships among them and other diabetes complications because the selected features can help us understand the causal relationship between features and classes.

Author Contributions: Conceptualization, A.S.-Y.L.; Data curation, J.-L.T. and W.-C.L.; Funding acquisition, A.S.-Y.L. and W.-C.L.; Investigation, Y.-D.J. and J.-S.H.; Methodology, A.S.-Y.L. Y.-D.J. and W.-C.L.; Project administration, A.S.-Y.L.; Validation, A.S.-Y.L., Y.-D.J. and W.-C.L.; Writing—original draft, A.S.-Y.L. and W.-C.L.; Writing—review & editing A.S.-Y.L. and W.-C.L. All authors have read and agreed to the published version of the manuscript.

References

1. Fritschi, C.; Quinn, L. Fatigue in patients with diabetes: a review. *J. Psychosom. Res.* **2010**, *69*, 33–41. [CrossRef] [PubMed]

2. Faulkner, J.A.; Brooks, S.V.; Zebra, E. Skeletal Muscle Weakness and Fatigue in Old Age: Underlying Mechanisms. In *Annual Review of Gerontology and Geriatrics: Special Focus on the Biology of Aging*; Cristofalo, V.J., Ed.; Springer Publishing: New York, NY, USA, 1990; Volume 10, pp. 147–166.

3. Fried, L.P.; Tangen, C.M.; Walston, J.; Newman, A.B.; Hirsch, C.; Gottdiener, J.; Seeman, T.; Tracy, R.; Kop, W.J.; Burke, G.; et al. Cardiovascular Health Study Collaborative Research Group: Frailty in older adults: evidence for a phenotype. *J. Gerontol. A Biol. Sci. Med. Sci.* **2001**, *56*, 146–156. [CrossRef] [PubMed]

4. American Diabetes Association. Economic costs of diabetes in the US in 2012. *Diabetes Care* **2013**, *36*, 1033–1046. [CrossRef]

5.　Ray, J.A.; Valentine, W.J.; Secnik, K.; Oglesby, A.K.; Cordony, A.; Gordois, A.; Davey, P.; Palmer, A.J. Review of the cost of diabetes complications in Australia, Canada, France, Germany, Italy and Spain. *Curr. Med Res. Opin.* **2005**, *21*, 1617–1629. [CrossRef] [PubMed]

6.　Seo, Y.M.; Hahm, J.R.; Kim, T.K.; Choi, W.H. Factors affecting fatigue in patients with type II diabetes mellitus in Korea. *Asian Nurs Res. (Korean Soc. Nurs Sci.)* **2015**, *9*, 60–64. [CrossRef] [PubMed]

7.　Reutrakul, S.; Thakkinstian, A.; Anothaisintawee, T.; Chontong, S.; Borel, A.-L.; Perfect, M.M.; Janovsky, C.C.; Kessler, R.; Schultes, B.; Harsch, I.A.; et al. Sleep characteristics in type 1 diabetes and associations with glycemic control: systematic review and meta-analysis. *Sleep Med.* **2016**, *23*, 26–45. [CrossRef]

8.　Goedendorp, M.M.; Tack, C.J.; Steggink, E.; Bloot, L.; Bazelmans, E.; Knoop, H. Chronic fatigue in type 1 diabetes: highly prevalent but not explained by hyperglycemia or glucose variability. *Diabetes Care* **2014**, *37*, 73–80. [CrossRef]

9.　Hamilton, M.T.; Hamilton, D.G.; Zderic, T.W. Role of low energy expenditure and sitting in obesity, metabolic syndrome, type 2 diabetes, and cardiovascular disease. *Diabetes* **2007**, *56*, 2655–2667. [CrossRef]

10.　Avlund, K. Fatigue in older adults:an early indicator of the aging process? *Aging Clin. Exp. Res.* **2011**, *22*, 100–115. [CrossRef]

11.　Bot, M.; Pouwer, F.; de Jonge, P.; Tack, C.; Geelhoed-Duijvestijn, P.; Snoek, F. Differential associations between depressive symptoms and glycaemic control in outpatients with diabetes. *Diabet. Med.* **2013**, *30*, e115–e122. [CrossRef]

12.　Eaton, W.W.; Armenian, H.; Gallo, J.; Pratt, L.; Ford, D.E. Depression and risk for onset of type II diabetes: a prospective population-based study. *Diabetes Care* **1996**, *19*, 1097–1102. [CrossRef] [PubMed]

13.　Fritschi, C.; Quinn, L.; Hacker, E.D.; Penckofer, S.M.; Wang, E.; Foreman, M.; Ferrans, C.E. Fatigue in women with type 2 diabetes. *Diabetes Educ.* **2012**, *38*, 662–672. [CrossRef] [PubMed]

14.　El_Jerjawi, N.S.; Abu-Naser, S.S. Diabetes Prediction Using Artificial Neural Network. *Int. J. Adv. Sci. Technol.* **2018**, *121*, 55–64.

15.　Eswari, T.; Sampath, P.; Lavanya, S. Predictive methodology for diabetic data analysis in big data. *Procedia Comput. Sci.* **2015**, *50*, 203–208.

16.　Jain, D.; Singh, V. Feature selection and classification systems for chronic disease prediction: A review. *Egypt. Inform. J.* **2018**, *19*, 179–189. [CrossRef]

17.　Ronksley, P.E.; Hemmelgarn, B.R.; Heitman, S.J.; Hanly, P.J.; Faris, P.D.; Quan, H.; Tsai, W.H. Obstructive sleep apnoea is associated with diabetes in sleepy subjects. *Thorax* **2009**, *64*, 834–839. [CrossRef]

18.　Cuellar, N.G.; Ratcliffe, S.J. A comparison of glycemic control, sleep, fatigue, and depression in type 2 diabetes with and without restless legs syndrome. *J. Clin. Sleep Med.* **2008**, *4*, 50–56. [CrossRef]

19.　Paschalides, C.; Wearden, A.J.; Dunkerley, R.; Bundy, C.; Davies, R.; Dickens, C.M. The associations of anxiety, depression and personal illness representations with glycaemic control and health-related quality of life in patients with type 2 diabetes mellitus. *J. Psychosom. Res.* **2004**, *57*, 557–564. [CrossRef]

20.　Reutens, A.T.; Atkins, R.C. Epidemiology of diabetic nephropathy. *Diabetes Kidney* **2011**, *170*, 1–7.

21.　Singh, R.; Teel, C.; Sabus, C.; McGinnis, P.; Kluding, P. Fatigue in type 2 diabetes: impact on quality of life and predictors. *PLoS ONE* **2016**, *11*, e0165652. [CrossRef]

22.　Gandhi, G.Y.; Murad, M.H.; Fujiyoshi, A.; Mullan, R.J.; Flynn, D.N.; Elamin, M.B.; Swiglo, B.A.; Isley, W.L.; Guyatt, G.H.; Montori, V.M. Patient-important outcomes in registered diabetes trials. *JAMA* **2008**, *299*, 2543–2549. [CrossRef] [PubMed]

23.　Ogurtsova, K.; da Rocha Fernandes, J.D.; Huang, Y.; Linnenkamp, U.; Guariguata, L.; Cho, N.H.; Cavan, D.; Shaw, J.E.; Makaroff, L.E. IDF Diabetes Atlas: Global estimates for the prevalence of diabetes for 2015 and 2040. *Diabetes Res. Clin. Pract.* **2017**, *128*, 40–50. [CrossRef] [PubMed]

24.　Tiesinga, L.J.; Dassen, T.W.; Halfens, R.J. Fatigue: a summary of the definitions, dimensions, and indicators. *Int. J. Nurs. Termin. Classif.* **1996**, *7*, 51–62. [CrossRef] [PubMed]

25.　Kumar, C.P. Application of Orem's Self-Care Deficit Theory and Standardized Nursing Languages in a Case Study of a Woman with Diabetes. *Int. J. Nurs. Terminol. Classif.* **2007**, *18*, 103–110. [CrossRef]

26.　Chandrashekar, G.; Sahin, F. A survey on feature selection methods. *Comput. Electr. Eng.* **2014**, *40*, 16–28. [CrossRef]

27.　Guyon, I.; Elisseeff, A. An introduction to variable and feature selection. *J. Mach. Learn. Res.* **2003**, *3*, 1157–1182.

28. Liu, H.; Yu, L. Toward integrating feature selection algorithms for classification and clustering. *IEEE Trans. Knowl. Data Eng.* **2005**, *17*, 491–502.

29. You, D.; Wu, X.; Shen, L.; He, Y.; Yuan, X.; Chen, Z.; Deng, S.; Ma, C. Online streaming feature selection via conditional independence. *Appl. Sci.* **2018**, *8*, 2548. [CrossRef]

30. Bolon-Canedo, V.; Sanchez-Marono, N.; Alonso-Betanzos, A. A review of feature selection methods on synthetic data. *Knowl. Inf. Syst.* **2013**, *34*, 483–519. [CrossRef]

31. Kumar, V.; Minz, S. Feature selection: a literature review. *Smart Comput. Rev.* **2014**, *4*, 211–229. [CrossRef]

32. Saeys, Y.; Inza, I.; Larranaga, P. A review of feature selection techniques in bioinformatics. *Bioinformatics* **2007**, *23*, 2507–2517. [CrossRef] [PubMed]

33. Tang, J.; Alelyani, S.; Liu, H. Feature selection for classification-a review. In *Data Classification Algorithms and Applications*; Aggarwal, C.C., Ed.; Chapman and Hall/CRC: Boca Raton, FL, USA, 2014.

34. Jolliffe, I.T. *Principal Component Analysis*; Springer Verlag: New York, NY, USA, 1986.

35. de la lglesia, B. Evolutionary computation for feature selection in classification problems. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2013**, *3*, 381–407. [CrossRef]

36. Xue, B.; Zhang, M.; Browne, W.N.; Yao, X. A survey on evolutionary computation approaches to feature selection. *IEEE Trans. Evol. Comput.* **2016**, *20*, 606–626. [CrossRef]

37. Xie, S.; Li, K.; Xiao, M.; Zhang, L.; Li, W. Key quality indicators prediction for web browsing with embedded filter feature selection. *Appl. Sci.* **2020**, *10*, 2141. [CrossRef]

38. Goldberg, D.E. *Genetic Algorithms in Search Optimization and Machine Learning*; Addition Wesley: Boston, MA, USA, 1989.

39. Maldonado, S.; López, J. Dealing with high-dimensional class-imbalanced datasets: Embedded feature selection for SVM classification. *Appl. Soft Comput.* **2018**, *67*, 94–105. [CrossRef]

40. Zhao, Z.; Liu, H. Spectral feature selection for supervised and unsupervised learning. In Proceedings of the 24th International Conference on Machine Learning, Corvalis Oregon, OR, USA, 20–24 June 2007; Ghahramani, Z., Ed.; ACM: New York, NY, USA, 2007; pp. 1151–1157.

41. Zhu, X.; Zhang, S.; Hu, R.; Zhu, Y.; Song, J. Local and global structure preservation for robust unsupervised spectral feature selection. *IEEE Trans. Knowl. Data Eng.* **2018**, *30*, 517–529. [CrossRef]

42. Tibshirani, R. Regression shrinkage and selection via the Lasso. *J. R. Stat. Soc. Ser. B (Methodological)* **1996**, *58*, 267–288. [CrossRef]

43. Bradley, P.S.; Mangasarian, O.L. Feature selection via concave minimization and support vector machines. In Proceedings of the 15th International Conference on Machine Learning, Madison, Wisconsin, WI, USA, 24–27 July 1998; Shavlik, J.W., Ed.; Morgan Kaufmann Publishers: San Francisco, CA, USA, 1998; pp. 82–90.

44. Zhu, Z.; Ong, Y.-S.; Dash, M. Wrapper-filter feature selection algorithm using a memetic framework. *IEEE Trans. Syst. Man, Cybern. Part B (Cybern.)* **2007**, *37*, 70–76. [CrossRef]

45. Quinlan, J.R. Induction of decision trees. *Mach. Learn.* **1986**, *1*, 81–106. [CrossRef]

46. Morchid, M.; Dufour, R.; Bousquet, P.-M.; Linares, G.; Torres-Moreno, J.-M. Feature selection using principal component analysis for massive retweet detection. *Pattern Recognit. Lett.* **2014**, *49*, 33–39. [CrossRef]

47. Mori, Y.; Lizuka, M.; Tarumi, T.; Tanaka, Y. Variable selection in principal component analysis. In *Statistical Methods for Biostatistics and Related Fields*; Hardle, W., Mori, Y., Vieu, P., Eds.; Springer: Berlin, Germany, 2007; pp. 265–283.

48. Kazemitabar, J.; Amini, A.; Bloniarz, A.; Talwalkar, A.S. Variable importance using decision trees. In Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017; pp. 425–434.

49. Liu, F.; Lee, D.-H.; Lagoa, R.; Kumar, S. EEG feature selection method based on decision tree. *Bio-Med Mater. Eng.* **2015**, *26*, S1019–S1025.

50. Tsai, C.-F.; Chen, Y.-C. The optimal combination of feature selection and data discretization: an empirical study. *Inf. Sci.* **2019**, *505*, 282–293. [CrossRef]

51. Liang, D.; Tsai, C.-F.; Wu, H.-T. The effect of feature selection on financial distress prediction. *Knowl. Based Syst.* **2015**, *73*, 289–297. [CrossRef]

52. Huang, M.-W.; Lin, W.-C.; Chen, C.-W.; Ke, S.-W.; Tsai, C.-F.; Eberle, W. Data preprocessing issues for incomplete medical datasets. *Expert Syst.* **2016**, *33*, 432–438. [CrossRef]

53. Cortes, C.; Vapnik, V.N. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [CrossRef]

54. Park, H.; Park, C.; Quinn, L.; Fritschi, C. Glucose control and fatigue in type 2 diabetes: the mediating roles of diabetes symptoms and distress. *J. Adv. Nurs.* **2015**, *71*, 1650–1660. [CrossRef]

55. Jain, A.; Sharma, R.; Choudhary, P.K.; Yadav, N.; Jain, G.; Maanju, M. Study of fatigue, depression, and associated factors in type 2 diabetes mellitus in industrial workers. *Ind. Psychiatry J.* **2015**, *24*, 179–184. [CrossRef]